Indag. Mathem., N.S., 12 (4), 489–503 December 17, 2001

Consecutive primes modulo 4

by Imre Z. Ruzsa*

Alfréd Rényi Mathematical Institute, Hungarian Academy of Sciences, Budapest, Pf. 127, H-1364 Hungary
e-mail: ruzsa@renyi.hu

Communicated by Prof. R. Tijdeman at the meeting of September 24, 2001

1. INTRODUCTION

Let $p_0 = 2, p_1 = 3, \dots$ be the sequence of primes. For $i \geq 1$ there are four possibilities for the pair of residues of p_i and p_{i+1} modulo 4. We use $\pi_{11}(x)$ to denote the number of pairs satisfying $p_i \equiv p_{i+1} \equiv 1 \pmod 4$, $p_{i+1} \leq x$. $\pi_{13}(x)$ etc. are defined analogously.

Since the number of primes in both residue classes $1 \pmod 4$ and $3 \pmod 4$ is infinite, we can immediately see that $\pi_{13}(x)$ and $\pi_{31}(x)$ tend to infinity; however, no comparably obvious reasoning is known for $\pi_{11}(x) \to \infty$. This follows from Littlewood's deep result asserting that $\pi(4,1,x) - \pi(4,3,x)$ is unbounded from both sides. Knapowski and Turán [3] gave a lower bound of the form $(\log x)^c$ for $\pi_{11}(x)$.

This result was greatly improved and extended by Shiu [7]. He proved, for arbitrary k, q and a satisfying $(a,q) = 1$, the existence of infinitely many i such that

$$p_i \equiv p_{i+1} \equiv \dots \equiv p_{i+k} \equiv a \pmod q.$$

Furthermore he proved that the number of such values of i with $p_i \leq x$ is $\gg x^{1-\varepsilon(x)}$, with a function $\varepsilon(x) \to 0$. His estimate for $\varepsilon(x)$ depends on k, q, a; in the case of pairs modulo 4, it is

*Supported by Hungarian National Foundation for Scientific Research, Grants No. T 025617 and T 29759

$$\varepsilon(x) = c\left(\frac{\log\log\log x}{\log\log x}\right)^{1/2}.$$

Our aim is to further improve this bound in the starting case, for residue classes modulo 4. Our estimate will be

$$\frac{x\log\log x}{(\log x)^2}.$$

Our method works in a slightly more general situation. We shall establish the following result.

Theorem. *Let $q \geq 3$ be an integer, and let A be a set of reduced residue classes modulo q satisfying $|A| = \phi(q)/2$. For $x > x_0(q)$ we have*

$$\#\{j : p_{j+1} \leq x, p_j \in A \pmod{q}, p_{j+1} \in A \pmod{q}\} \geq c_0 \frac{\phi(q)}{q^3} \frac{x\log\log x}{(\log x)^2},$$

with an absolute constant $c_0 > 0$.

Our method, unlike Shiu's, does not work for $|A| < \phi(q)/2$.

We mention some natural particular cases.

Corollary 1.1. *Let $q \geq 3$ be an integer, let χ be a nonprincipal real character modulo q and let $\delta \in \{1, -1\}$. We have*

$$\#\{j : p_{j+1} \leq x, \chi(p_j) = \chi(p_{j+1}) = \delta\} \geq c_0 \frac{\phi(q)}{q^3} \frac{x\log\log x}{(\log x)^2},$$

where the implicit constant depends on q.

Corollary 1.2. *Let q be 4 or 6 and let $\delta \in \{1, -1\}$. We have*

$$\#\{j : p_{j+1} \leq x,\ p_j \equiv p_{j+1} \equiv \delta \mod q\} \gg \frac{x\log\log x}{(\log x)^2}.$$

My proof, as well as Shiu's, applies an averaging argument that was invented by Maier and applied by him to establish irregularity properties of primes [4].

The Theorem will be proved in sections 2-3. In Section 4 we discuss some conjectures about the behaviour of π_{ij}, with a conditional result and numerical evidence.

Presently I cannot give any nontrivial lower bound for π_{13}. One could imagine an estimate along the following lines. There are many results that say that for some h, all, or almost all intervals of length h up to x contain a prime. Presumably some of these can be modified to show that almost all such intervals contain primes from both residue classes 1 and 3 (mod 4. In this way we could get $\pi_{13}(x) \gg x/h$. However, even if we could so adapt the best available estimates, we would be unable to obtain a bound of the form $x^{1+o(1)}$.

Here we prove the Theorem, save a lemma which is deferred to the next section.

In the sequel we consider q to be fixed, and the estimates will hold under the assumption that $x > x_0(q)$, without further mentioning. We will not calculate this bound $x_0(q)$. The depencence on q can probably be improved in the theorem, and the method also works for q tending to infinity sufficiently slowly; but the general result seems to be less interesing anyway than the particular cases described in Corollary 1.2, where q is at most 6.

Define

$$\chi(n) = \begin{cases} 1 & \text{if } n \in A \pmod q, \\ -1 & \text{if } n \notin A \pmod q, (n,q) = 1, \\ 0 & \text{if } (n,q) > 1. \end{cases}$$

(χ will not be assumed to possess any multiplicativity property.)

We will take an integer m coprime to q and consider (for suitable fixed T and variable n) the sum

$$f(n) = \sum_{n+1 \le j \le n+T, (j,m)=1} \chi(j).$$

Write $M = qm$. A large value of $f(n)$ signifies that in any interval of type

$$[kM + n + 1, kM + n + T] \tag{2.1}$$

among the integers coprime to M there is a preponderance of numbers with $\chi(j) = 1$. Then we expect such an imbalance for the primes as well, which will lead to consecutive primes with residue in A. We will achieve this by an averaging argument.

Let u_k and v_k denote the number of primes in an interval of type (2.1) satisfying $\chi(p) = 1$ or -1, respectively. Clearly the number of those primes p_j for which

$$kM + n + 1 \le p_j < p_{j+1} \le kM + n + T, \quad \chi(p_j) = \chi(p_{j+1}) = 1 \tag{2.2}$$

is at least $u_k - v_k - 1$, unless there is a prime with $\chi(p) = 0$. We will exclude this possibility by assuming $k \ge 1$, whence the primes considered satisfy $p > M \ge q$. As we are interesed in primes $\le x$, we restrict the range of k to satisfy $kM + n + T \le x$. We will use $n = 1, \dots, M$, hence a suitable upper bound is

$$K = \left[\frac{x - T}{M}\right] - 1.$$

In the union of these intervals we find at least

$$\sum_{k=1}^{K} (u_k - v_k - 1) \ge \sum (u_k - v_k) - \frac{x}{M}$$

such pairs. On the other hand we have

$$\sum(u_k - v_k) = \sum_{k=1}^{K} \sum_{kM+n+1 \le p \le kM+n+T} \chi(p) =$$

$$\sum_{j=n+1}^{n+T} \chi(j)\bigl(\pi(KM+j, M, j) - \pi(j, M, j)\bigr).$$

With the above choice of K we have clearly

$$\pi(x, M, j) - 3 \le \pi(KM+j, M, j) - \pi(j, M, j) \le \pi(x, M, j).$$

We summarize the above argument as follows. With

$$Z_n = \#\{j : p_{j+1} \le x, \quad (2.2) \text{ holds for some } k\}$$

we have

$$Z_n \ge \sum_{j=n+1}^{n+T} \chi(j)\pi(x, M, j) - \frac{x}{M} - 3T. \tag{2.3}$$

To estimate this we employ the prime number theorem for arithmetic progressions in its classical form (see e.g. [3]).

Lemma 2.1. *With a suitable absolute constant c we have*

$$\pi(x, M, j) = \frac{\operatorname{li} x}{\phi(q)} + O(x \exp -c\sqrt{\log x})$$

for all $M \le \exp c\sqrt{\log x}$, except possibly the multiples of a certain integer q_0, depending on x. This integer q_0 (if exists) satisfies $q_0 \gg (\log x)^{1-\varepsilon}$

The estimate of q_0 can be improved applying Siegel's theorem, but all we need here is that $q_0 \to \infty$ as $x \to \infty$.

Now we specify m. Let p_0 be a prime whose exponent in q_0 is larger than in q (typically one that divides q_0 but not q); such a prime exists as soon as $q_0 > q$. We put

$$Q = (\log x)^{1/3}, \quad m = \prod_{p < Q, p \nmid 2qp_0} p, \quad M = qm. \tag{2.4}$$

This ensures $q_0 \nmid M$. An application of the above lemma to (2.3) yields

$$Z_n > f(n)\frac{\operatorname{li} x}{\phi(M)} - O\bigl(Tx \exp -c\sqrt{\log x}\bigr) - \frac{x}{M} - 3T > f(n)\frac{\operatorname{li} x}{\phi(M)} - \frac{2}{M}x$$

since our M will be $< \exp \varepsilon\sqrt{\log x}$ and we will choose $T \ll (\log x)^2$.

We sum this inequality for those $1 \le n \le M$ for which $f(n) > 0$, and use the trivial $Z_n \ge 0$ otherwise. We obtain

$$\sum Z_n > \frac{\operatorname{li} x}{\phi(M)} \sum f(n)^+ - 2x, \tag{2.5}$$

where $a^+ = \max(a, 0)$.

We estimate this sum over f as follows.

Lemma 2.2. *With M defined as in (2.4), uniformly in the range* $\log x \le T \le (\log x)^3$ *we have*

$$\sum_{n=1}^{M} f(n)^+ \ge c_1 \sqrt{\frac{\phi(q)}{q^3} TM\phi(M)} \tag{2.6}$$

with a certain positive absolute constant c_1.

The proof of this lemma is deferred to the next section.

A combination of (2.5) and (2.6) gives

$$\sum Z_n > c_1 \sqrt{\frac{\phi(q)}{q^3} \frac{TM}{\phi(M)}} \operatorname{li} x - 2x > c_1 \sqrt{\frac{\phi(q)}{q^3} \frac{TM}{\phi(M)}} \frac{x}{\log x} - 2x,$$

taking into account that $\operatorname{li} x > x/\log x$.

Now $\sum Z_n$ counts pairs of consecutive primes with residues in A, and each such pair is counted at most T times. Hence we obtain

$$\# \text{ of suitable prime pairs } \ge \frac{1}{T} \sum Z_n > c_1 \sqrt{\frac{\phi(q)}{q^3} \frac{M}{T\phi(M)}} \frac{x}{\log x} - \frac{2x}{T}.$$

The theorem follows by setting

$$T = \frac{16q^3}{c_1^2 \phi(q)} \frac{\phi(M)}{M} (\log x)^2$$

and observing that

$$\frac{\phi(M)}{M} = \prod_{2<p<Q, p \ne p_0} \left(1 - \frac{1}{p}\right) \ll \frac{1}{\log Q} \ll \frac{1}{\log \log x}.$$

3. PROOF OF LEMMA 2.2

We are going to estimate the first, second and fourth moments of f.

We shall consider f as a function on $\mathbb{Z}_M$, the ring of residues modulo M, which we shall identify with $\mathbb{Z}_q \times \mathbb{Z}_m$ and write $n = \{u, v\}$, where u is a residue modulo q and v is one modulo m. We write $\gamma(v) = 1$ if $(v, m) = 1$, 0 otherwise. With this notation we have

$$f(n) = f(u, v) = \sum_{i=1}^{T} \chi(u+i)\gamma(v+i).$$

Lemma 3.1.

$$\sum_{n=1}^{M} f(n) = 0.$$

Proof. Clearly we have for each i

$$\sum_{u,v} \chi(u+i)\gamma(v+i) = \sum \chi(u)\gamma(v) = \left(\sum \chi(u)\right)\left(\sum \gamma(v)\right) = 0$$

by the assumption $|A| = \phi(q)/2$. $\square$

Lemma 3.2. *For $x > x_0$ and $T \leq (\log x)^3$ we have*

$$\sum_{d|m, d>T} \frac{1}{d} \gg \frac{m}{\phi(m)}.$$

Proof. Recall that

$$M = \prod_{2<p<Q, p\neq p_0} p, \qquad Q = (\log x)^{1/3}.$$

Take a number $\alpha \in (0,1)$ and a positive integer k. Define $Q_i = Q^{\alpha^i}$. We shall consider only divisors of M in the form $d = p_1 p_2 \ldots p_k d'$, where $p_i \in [Q_i, Q_{i-1})$ and every prime divisor of d' is $< Q_k$. These divisors are all

$$> Q_1 Q_2 \ldots Q_k = Q^{\alpha + a^2 + \ldots \alpha^k} \geq Q^9 = (\log x)^3,$$

provided α and k are chosen so that $\alpha + a^2 + \ldots \alpha^k \geq 9$. Possible values are, say, $\alpha = 0.95$ and $k = 13$.

The sum of reciprocals of these divisors is

$$\prod_{p<Q_k} \left(1 + \frac{1}{p}\right) \prod_{i=1}^{k} \sum_{Q_i \leq p < Q_{i-1}} \frac{1}{p}. \tag{3.1}$$

By the familiar formula on the sum of reciprocals of primes we have

$$\sum_{Q_i \leq p < Q_{i-1}} \frac{1}{p} = \log\log Q_{i-1} - \log\log Q_i + o(1) \to \log\frac{1}{\alpha}.$$

The first product in (3.1) can be written as

$$\frac{m}{\phi(m)} \prod_{p<Q_k} \left(1 - \frac{1}{p^2}\right) \prod_{Q_k \leq p < Q} \left(1 - \frac{1}{p}\right).$$

Here the first product is larger than the corresponding product for all primes, which has value $6/\pi^2$. The second product is asymptotically equal to

$$\frac{\log Q_k}{\log Q} = \alpha^k$$

by Mertens' theorem. Therefore we have

$$\sum_{d|m, d>T} \frac{1}{d} \geq \beta \frac{m}{\phi(m)}$$

with any

$$\beta < \frac{6}{\pi^2} \left(\alpha \log\frac{1}{\alpha}\right)^k.$$

for sufficiently large x. $\square$

Lemma 3.3. *For $T \leq (\log x)^3$ we have*

$$\sum_{n=1}^{M} f(n)^2 \gg T\phi(m)\phi(q).$$

Proof. After a term-by-term multiplication and collecting the identical products we obtain

$$\begin{aligned} S &= \sum_{n=1}^{M} f(n)^2 \\ &= \sum_{u_1, u_2 \in \mathbb{Z}_q} \sum_{v_1, v_2 \in \mathbb{Z}_m} \chi(u_1)\chi(u_2)\gamma(v_1)\gamma(v_2)\delta(u_1 - u_2, v_1 - v_2), \end{aligned}$$

where $\delta(u, v)$ is the number of pairs of integers i, j such that $1 \leq i, j \leq T$ and $i - j \equiv u \pmod q), \equiv v \pmod m$. On introducing the new variables $u = u_2 - u_1$, $v = v_2 - v_1$ we can write this as

$$S = \sum_{u, u_1, v, v_1} \chi(u_1)\chi(u_1 + u)\gamma(v_1)\gamma(v_1 + v)\delta(u, v). \tag{3.2}$$

Now we calculate the sum over v_1, with the other three variables fixed.

$$\sum_{v_1} \gamma(v_1)\gamma(v_1 + v)$$

is the number of residues v_1 modulo m such that both v_1 and $v_1 + v$ are coprime to m; by an obvious sieve we have

$$\begin{aligned} \sum_{v_1} \gamma(v_1)\gamma(v_1 + v) &= m \prod_{p|m, p\nmid v} \left(1 - \frac{2}{p}\right) \prod_{p|m, p|v} \left(1 - \frac{1}{p}\right) \\ &= m \prod_{p|m} \left(1 - \frac{2}{p}\right) \prod_{p|m, p|v} \frac{p-1}{p-2}. \end{aligned}$$

By introducing the function $H(n) = \prod_{p|n} (p - 2)$ we can write the first product as $H(m)/m$ and the second as $\sum_{d|(v,m)} 1/H(d)$. (Recall that m is odd by definition.) Hence

$$\sum_{v_1} \gamma(v_1)\gamma(v_1 + v) = H(m) \sum_{d|(v,m)} \frac{1}{H(d)}.$$

On substituting this into (3.2) we obtain

$$\begin{aligned} S &= H(m) \sum_{u, u_1, v} \chi(u_1)\chi(u_1 + u)\delta(u, v) \sum_{d|(v,m)} \frac{1}{H(d)} \\ &= H(m) \sum_{d|m} \frac{1}{H(d)} \left(\sum_{u, u_1} \chi(u_1)\chi(u_1 + u) \sum_{j=0}^{m/d - 1} \delta(u, dj) \right). \end{aligned}$$

The inner sum is always nonnegative, since it is equal to

$$\frac{1}{m}\sum_{j=1}^{d}\left(\sum_{n} g_{dj}(n)\right)^2,$$

where

$$g_{dj}(n) = \sum_{1\le i\le T,\, n+i\equiv j \pmod d} \chi(n+i).$$

To give a lower estimate we keep only the terms corresponding to $d > T$. In this case the only nonzero terms in the inner sum correspond to $u = 0, j = 0$ and we obtain

$$\begin{aligned} S &\ge H(m) \sum_{d|m,\, d>T} \frac{1}{H(d)} \sum_{u_1} \chi(u_1)^2 \delta(0,0) \\ &= H(m)\phi(q)T \sum_{d|m,\, d>T} \frac{1}{H(d)} \\ &\ge H(m)\phi(q)T \sum_{d|m,\, d>T} \frac{1}{d}. \end{aligned}$$

To estimate this we observe that

$$\frac{H(m)}{m} = \left(\frac{\phi(m)}{m}\right)^2 \prod_{p|m}\left(1-\frac{1}{(p-1)^2}\right) > \left(\frac{\phi(m)}{m}\right)^2 \prod_{p\ge 3}\left(1-\frac{1}{(p-1)^2}\right).$$

An appeal to the previous lemma completes the proof. □

Lemma 3.4. *For $T \ge q\big(1 + m/\phi(m)\big)$ we have*

$$\sum_{n=1}^{M} f(n)^4 \ll T^2 \frac{\phi(m)^2}{m} q^2 \phi(q).$$

Proof. We have

$$f(n) = \sum_{i=1}^{q} f_i(n) + f^*(n),$$

where

$$f_i(n) = \sum_{1\le j\le T,\, j\equiv i \pmod q} \chi(n+j)\left(\gamma(n+j) - \frac{\phi(m)}{m}\right)$$

and

$$f^*(n) = \sum_{j=1}^{T} \chi(n+j)\frac{\phi(m)}{m}.$$

We shall estimate the fourth moments of the functions f_i and f^*.

By the periodicity of χ we see that $|f^*(n)| \le q\phi(m)/m$, hence

$$\sum_{n} f^*(n)^4 \le q^4\phi(m)^4 m^{-4} M = q^5 \phi(m)^4 m^{-3}. \tag{3.3}$$

Now we show that

$$\sum_n f_i(n)^4 \ll \phi(q)\frac{\phi(m)^2}{m}L^2. \tag{3.4}$$

We have (for $n = \{u, v\}$)

$$f_i(n) = \chi(u+i)\sum_{l=0}^{L}\left(\gamma(v+i+lq) - \frac{\phi(m)}{m}\right),$$

where

$$L = \left[\frac{T-i}{q}\right].$$

Hence

$$\begin{aligned}\sum f_i(n)^4 &= \sum_u \chi(u+i)^4 \sum_v \left(\sum_{l=0}^{L}\left(\gamma(v+i+lq) - \frac{\phi(m)}{m}\right)\right)^4\\ &= \phi(q)\sum_v \left(\sum_{l=0}^{L}\left(\gamma(v+i+lq) - \frac{\phi(m)}{m}\right)\right)^4.\end{aligned}$$

Since $(q, m) = 1$, we can find a q^* such that $qq^* \equiv 1 \pmod{m}$ and we have

$$\gamma(v+i+lq) = \gamma(q^*(v+i)+l).$$

As v runs over all residues modulo m, so does $w = q^*(v+i)$, thus we have

$$\sum_v \left(\sum_l\right)^4 = \sum_w \left(\sum_{l=0}^{L}\left(\gamma(w+l) - \frac{\phi(m)}{m}\right)\right)^4.$$

This sum is known to be

$$\ll m\left(\frac{\phi(m)}{m}L\right)^2$$

as soon as $L > m/\phi(m)$, by a theorem of Montgomery and Vaughan [6] (see also in [5]). This assumption follows from the assumption of the lemma on the size of T. Thus (3.4) is proved.

By the power-mean inequality we have

$$\begin{aligned}\sum f(n)^4 &\le (q+1)^3 \sum_n \left(f^*(n)^4 + \sum_i f_i(n)^4\right)\\ &\ll q^4\phi(q)\frac{\phi(m)^2}{m}L^2 + q^8\phi(m)^4 m^{-3}\\ &\ll q^2\phi(q)T^2\frac{\phi(m)^2}{m}.\end{aligned}$$

In the last inequality we used the definition of L, and to see that the second term is of smaller order of magnitude than the first we recall that $m/\phi(m) \to \infty$ as $x \to \infty$. This concludes the proof of Lemma 3.4. $\square$

Proof of Lemma 2.2.

By Hölder's inequality and the last two lemmas we have

$$\sum |f(n)| \geq \left(\sum f(n)^2\right)^{3/2} \left(\sum f(n)^4\right)^{-1/2} \gg \frac{\phi(q)}{q} \sqrt{Tm\phi(m)}.$$

Thus by Lemma 3.1

$$\sum f(n)^+ = \sum \frac{f(n) + |f(n)|}{2} = \frac{1}{2} \sum |f(n)| \gg \frac{\phi(q)}{q} \sqrt{Tm\phi(m)}.$$

To obtain the Lemma we now substitute $m = M/q$, $\phi(m) = \phi(M)/\phi(q)$ and observe that $\phi(M) \gg M/\log\log m > M/\log\log x$. $\square$

4. ON THE KNAPOWSKI-TURÁN CONJECTURE

Knapowski and Turán [3] conjecture that

$$\pi_{11}(x) = o\left(\frac{x}{\log x}\right).$$

Erdős [1] on commenting the above paper writes he would rather think that this is $\sim (1/4)x/\log x$. Here we show that Knapowski and Turán's conjecture contradicts certain hypotheses that are generally hold plausible. Presently I cannot decide the truth of Erdős's conjecture even on the basis of unproved hypotheses.

Let $T(d, x)$ denote the number of those primes $p \leq x$ for which $p + d$ is also prime. There is a well-known conjecture, due to Hardy and Littlewood, about the asymptotic behaviour or $T(2, x)$, and the same heuristic yields the conjecture that

$$T(d, x) = \prod_{p|d} \left(1 + \frac{1}{p-1}\right) \prod_{p \nmid d} \left(1 - \frac{1}{(p-1)^2}\right) \frac{x}{(\log x)^2} + o\left(\frac{x}{(\log x)^2}\right).$$

We need less than this in the sense that we need only a lower bound; on the other hand we need it uniformly for $d < c \log x$. Observe that the coefficient of $x/(\log x)^2$ in the above formula is

$$\geq \prod_{p>2} \left(1 - \frac{1}{(p-1)^2}\right) > 0$$

for every even d.

Statement 4.1. *Assume that there is a constant $c > 0$ with the following property: for all sufficiently large x and even $d < c \log x$ we have*

$$(4.1) \qquad T(d, x) > c\frac{x}{(\log x)^2}.$$

Then there is a $c' > 0$ such that

$$\pi_{ij}(x) > c' \frac{x}{\log x}$$

for all sufficiently large x and each choice of $i, j \in \{1, -1\}$.

The proof is based on the simple observation that if two primes are close, they are likely to be consecutive. This will be used in the following form.

Lemma 4.2. *There is an absolute constant C with the following property. For every x and y we have*

$$\#\{p_1 < p_2 < p_3 \le p_1 + y, p_1 \le x, p_1, p_2, p_3 \ \text{ primes }\} \le Cxy^2(\log x)^{-3}. \tag{4.2}$$

Proof. Since this is a rather standard argument, we suppress some routine details.

First, for given integers $0 < a < b$, by Brun's or Selberg's sieve we obtain

$$\#\{p \le x : p,\, p+a,\, p+b \ \text{ primes}\} \le C_1 x \prod_{p \le x} \left(1 - \frac{\varrho(p)}{p}\right),$$

where $\varrho(p)$ is the number of different residues modulo p among the numbers $0, a, b$. The value of $\varrho(p)$ is typically 3; it is 2 if p divides a, b or $a-b$, and it is 0 if $p|(a,b)$.

To estimate this observe that

$$1 - \frac{\varrho(p)}{p} \le \left(1 + \frac{1}{p}\right)^{-\varrho(p)}$$

and recall that

$$\prod_{p \le x} 1 + \frac{1}{p} \ge C_2 \log x.$$

We obtain

$$\#\{p \le x : p,\, p+a,\, p+b \ \text{ primes}\} \le C_3 x (\log x)^{-3} h\big(ab(b-a)\big) h\big((a,b)\big)$$

with $h(a) = \prod_{p|a} (1 + 1/p)$.

We shall need that

$$\sum_{n \le N} h(n)^3 \le C_4 N. \tag{4.3}$$

To show this, we represent the summand in the form

$$h(n)^3 = \sum_{d|n} g(d).$$

Then the sum in (4.3) is

$$= \sum g(d) \left[\frac{N}{d}\right] \le N \sum_{d=1}^{\infty} \frac{|g(d)|}{d} \le C_4 N.$$

To see the convergence of the last sum one proceeds in the usual way, by developing it as an Euler product.

To prove (4.2) we need to show that

$$\sum_{1 \le a < b \le y} h\big(ab(b-a)\big)h\big((a,b)\big) \le C_5 y^2. \tag{4.4}$$

To deduce (4.4), write $(a,b) = d$, $a = du$, $b = dv$. The sum in (4.4) is

$$\le \sum_{d \le y} h(d)^2 \sum_{1 \le u < v \le y/d} h(u)h(v)h(v-u). \tag{4.5}$$

To estimate the inner sum we apply the inequality of the geometric mean to get

$$\begin{aligned} \sum_{1 \le u < v \le z} h(u)h(v)h(v-u) &\le \sum_{1 \le u < v \le z} \frac{1}{3}\Big(h(u)^3 + h(v)^3 + h(v-u)^3\Big) \\ &\le z \sum_{u \le z} h(u)^3 \le C_4 z^2 \end{aligned}$$

by (4.3). Thus (4.5) can be estimated as

$$\le C_4 y^2 \sum_{d \le y} \frac{h(d)^2}{d^2} \le C_5 y^2.$$

In the last step we need the convergence of the sum

$$\sum_{d \le 1}^{\infty} \frac{h(d)^2}{d^2},$$

which again follows from the Euler product representation.

This concludes the proof of Lemma 4.2. $\square$

Proof of Statement 4.1. We will select a number $0 < \alpha < c$ and sum our assumed lower estimate (4.1) for $0 < d < \alpha \log x$ separately for $d \equiv 0 \pmod 4$ and $d \equiv 2 \pmod 4$. We obtain, for $a = 0$ or 2, that

$$\sum_{d < \alpha \log x,\, d \equiv a \ (\mathrm{mod}\ 4)} T(d,a) \ge c_1 \alpha \frac{x}{\log x}, \tag{4.6}$$

say with $c_1 = c/5$.

This sum counts the number of prime pairs $p < p' < p + \alpha \log x$ satisfying $p' - p \equiv a \pmod 4$. Some of these pairs are consecutive and some are not. By the Lemma above, the number of nonconsecutive pairs is at most $C\alpha^2 x/\log x$. If α is such that $C\alpha^2 \le c_1\alpha/2$, say $\alpha = \min\big(c, c_1/(2C)\big)$, then (4.6) yields

$$\#\{j : p_{j+1} \le x, p_{j+1} - p_j \equiv a \ (\mathrm{mod}\ 4)\} \ge c_2 \frac{x}{\log x}.$$

We can rewrite this, for $a = 0$ and $a = 2$, as

$$\pi_{11}(x) + \pi_{33}(x) \ge c_2 \frac{x}{\log x} \tag{4.7}$$

and

$$(4.8) \qquad \pi_{13}(x) + \pi_{31}(x) \geq c_2 \frac{x}{\log x},$$

respectively.

Now to get a bound for the individual terms we just need to observe that

$$(4.9) \qquad \pi_{13}(x) - \pi_{31}(x) = O(1),$$

which follows immediately from the definition, and

$$(4.10) \qquad \pi_{11}(x) - \pi_{33}(x) = o\left(\frac{x}{\log x}\right),$$

which can be deduced from the prime number of arithmetic progressions as follows. We know that

$$\pi(x, 4, 1) = \pi_{11}(x) + \pi_{13}(x) + O(1) = \frac{\operatorname{li} x}{2} + o\left(\frac{x}{\log x}\right),$$

and similarly

$$\pi(x, 4, 3) = \pi_{31}(x) + \pi_{33}(x) + O(1) = \frac{\operatorname{li} x}{2} + o\left(\frac{x}{\log x}\right).$$

By substracting these equations and applying (4.9) we obtain (4.10).

By taking the sum and difference of formulas (4.7) and (4.10), then of formulas (4.8) and (4.9) we obtain the asserted bounds for all functions π_{ij}. $\square$

Finally we present some numerical results. The above table lists the distribution of consecutive pairs of pseudoprimes in the four possible combinations of residue classes modulo 4, in some intervals of type $[10^k, 10^k + y]$, for every $5 \leq k < 20$ and every even $20 \leq k \leq 80$. For this calculation (by practical reasons) a pseudoprime is an integer p free of prime factors $< 2^{17}$ and satisfying

$$2^{(p-1)/2} \equiv \left(\frac{2}{p}\right) \bmod p.$$

The value of y is selected so that the interval contains 20001 pseudoprimes, that is, 20000 pairs. It is possible that some data below come from proper pseudoprimes, but also it is likely that the error induced is less than the 'noise' due to the selection of comparatively short intervals.

The first column gives the value of k, the next four the statistics for each pair of residues (the sum of these is 20000), the last one the negative of the correlation, in percents. If the Knapowski-Turán conjecture is true, this number should tend to 100 (on average), under the Erdős conjecture it should tend to 0.

k	(1, 1)	(1, 3)	(3, 1)	(3, 3)	− correlation in percents
5	4208	5766	5767	4259	15.330750
6	4351	5672	5673	4304	13.450626
7	4390	5598	5599	4413	11.970148
8	4431	5532	5533	4504	10.651474
9	4517	5449	5449	4585	8.981260
10	4609	5432	5432	4527	8.641826
11	4647	5405	5406	4542	8.112980
12	4654	5398	5399	4549	7.972976
13	4795	5303	5303	4599	6.070187
14	4722	5266	5266	4746	5.320152
15	4686	5308	5308	4698	6.160038
16	4666	5303	5303	4728	6.061019
17	4724	5298	5299	4679	5.970536
18	4774	5233	5234	4759	4.670059
19	4709	5255	5256	4780	5.111324
20	4765	5204	5204	4827	4.081000
22	4804	5236	5235	4725	4.711634
24	4758	5217	5217	4808	4.340652
26	4795	5213	5214	4778	4.270075
28	4807	5174	5174	4845	3.480374
30	4738	5209	5208	4845	4.172981
32	4784	5224	5223	4769	4.470059
34	4894	5165	5165	4776	3.303596
36	4986	5116	5116	4782	2.330646
38	4769	5193	5193	4845	3.861500
40	4929	5125	5124	4822	2.492933
42	4842	5122	5122	4914	2.441328
44	4860	5175	5175	4790	3.501268
46	4873	5108	5107	4912	2.150388
48	4801	5101	5100	4998	2.019898
50	4972	5072	5073	4883	1.452009
52	4901	5095	5096	4908	1.910012
54	4811	5170	5171	4848	3.410354
56	4949	5098	5099	4854	1.972301
58	5037	5121	5121	4721	2.445575
60	4861	5064	5065	5010	1.295622
62	4876	5100	5101	4923	2.010563
64	4818	5082	5081	5019	1.640266
66	5061	5020	5020	4899	0.406588
68	4951	5104	5104	4841	2.083088
70	4869	5053	5053	5025	1.066149
72	4918	5099	5099	4884	1.980295
74	4919	5115	5115	4851	2.301183
76	5003	5053	5053	4891	1.063169
78	4872	5116	5116	4896	2.320147
80	4986	5106	5105	4803	2.118549

ACKNOWLEDGEMENT

I am grateful to Daniel K.I. Shiu for showing me his paper before publication, and to a referee for several corrections.

REFERENCES

[1] Erdős, P. – Personal reminiscences of the work of Paul Turán. In: Collected Papers of Paul Turán, ed. P. Erdős. Akadémiai Kiadó, Budapest 1990, vol. 1, XIV–XX (1990).

[2] Karatsuba, A.A. – Basic analytic number theory. Springer-Verlag, New York-Berlin (1993).

[3] Knapowski, S. and Turán, P. – On prime numbers $\equiv 1$ resp. 3 mod 4. In: Number Theory and Algebra. Academic Press, New York, 157–165 (1977). Reprinted in: Collected Papers of Paul Turán, ed. P. Erdős. Akadémiai Kiadó, Budapest 1990, vol. 3, p. 2534–2542.

[4] Maier, H. – Chains of large gaps between consecutive primes. Adv. Math. **39**, 257–269 (1981)

[5] Montgomery, H.L. – Ten lectures on the interface between analytic number theory and harmonic analysis. Regional Conf. Ser. Math. **84**, Amer. Math. Soc., Providence, R.I. (1994).

[6] Montgomery, H.L. and Vaughan, R.C. – On the distribution of reduced residues. Ann. Math. **123**, 311–333 (1986).

[7] Shiu, D.K.I. – Strings of conruent primes. J. London Math. Soc. **61**, 359–373 (2000).

(Received March 2001)